

MEMBRANE PENETRATING PEPTIDES AND USES THEREOF

This application claims the benefit of U.S. Provisional Application No. 60/227,647, filed August 25, 2001 and GB Application 0103110.3, filed February 7, 2001.

FIELD OF THE INVENTION

The invention relates to membrane penetrating peptides useful as *in vitro*, *ex vivo* and *in vivo* delivery devices for intracellular delivery of a compound of interest to cells *in vitro*, *ex vivo* and *in vivo*, compositions comprising the same and methods of using the same. The invention also includes identification of additional membrane penetrating peptides useful as delivery devices for intracellular delivery of a compound of interest to cells *in vitro*, *ex vivo* and *in vivo*.

BACKGROUND OF THE INVENTION

The delivery of small molecules, oligonucleotides, and proteins through biological membranes is a major challenge facing therapy and validation paradigms. It has recently been established that transducing peptides derived from Antennapedia, TAT-HIV, and VP22 can penetrate biological membranes, act as cargo vehicles, and target to specific subcellular compartments. Here we show the identification of a nuclear localization sequence (NLS) within human Period 1 (hPER1) circadian protein that functions as a transducing peptide. More importantly, using database mining, we have uncovered additional transducing peptides embedded within the NLS's of other proteins and extend the number of gene-encoded transducing peptides from 3 to 14. Our data suggest that transducing peptides are found within NLS's and are prevalent, diverse, and distributed widely throughout the genome. It is well established that certain extracellular and intracellular proteins are targeted to specific organelles within a cell, transmembrane or secreted from the cell. The biological mechanisms by which intracellular protein targeting occurs continues to be characterized, but is well recognized that one mechanism for localization occurs by virtue of specific leader sequence contained within the protein of interest, or intraprotein sequence. Localization of proteins within selected cellular organelles is aided by specific targeting sequences. A number of nuclear localization sequences (NLSs) have been identified in proteins that permit the protein to be transported or otherwise pass from the cytoplasm into the nuclear membrane.

Fusion proteins containing the targeting sequence and another, otherwise non-targeted protein, are localized in the selected cellular organelle depending on the targeting sequence selected. For example, Ferullo, J. M. and Paget, E. FR 279695, disclose selective compartmentalization of an hydroxyphenylpyruvate dioxygenase (HPPD) fused to a signal sequence directing the enzyme to a cellular compartment other than the cytosol, e.g., a vacuole. Similarly, WO 0147950 (Wehrle-Haller, Bernhard M.; Imhof, Beat A) identify a new determinant responsible for basolateral targeting and prolonged exposure of cell-surface-anchored growth factors at cell surfaces. The signal is a mono-leucine dependent basolateral sorting signal consisting of the amino acid sequence X1h2X3h4Lp5p6, wherein: X1 represents a polar amino acid residue or alanine, h2 represents any hydrophobic amino acid residue, X3 represents any amino acid residue, h4 represents any hydrophobic amino acid residue, except leucine and isoleucine, L represents a leucine residue, p5 represents any polar amino acid residue, and p6 represents any polar amino acid. Richardson, A. E., et al., Plant J. (2001), 25(6), 641-649 describe manipulation of the enzyme aspergillus phytase to include the signal peptide sequence from the carrot extensin gene. The resulting fusion protein was only effective when secreted as an extracellular enzyme into the adjacent soil, and resulted in a 20-fold increase in total root phytase activity in transgenic lines and subsequent improved phosphorus nutrition, such that the growth and phosphorus content of the plants was equivalent to control plants supplied with inorganic phosphate. WO 0132894 (Lok, S.) disclose use of the signal anchor domain sequences of type II cell surface proteins to anchor recombinant proteins into surface of transfected cells. A characteristic feature of type II cell surface proteins is that they are held within the cellular membrane by a single hydrophobic transmembrane domain and are oriented with their C-terminus outside the cell.

More recently, a few proteins have been identified which are capable of passing through the cellular membrane without requiring active transport mechanisms or 'pores'. It is recently established that membrane penetrating peptides (MPPs, also known as protein transduction domain, "PTD") derived from Antennapedia, TAT, and VP22 can penetrate biological membranes and target to specific subcellular compartments. None of these previously disclosed proteins are derived from mammalian proteins. The present invention is directed to the discovery that polypeptides derived from mammalian or yeast proteins nuclear localization sequences (NLSs) or overlapping with NLS's are capable of acting as MPPs, and identification of a specific polypeptide sequences capable of penetrating cellular membranes, even when conjugated to large proteins, such as biologically active proteins, or other organic compounds.

Nuclear transport is essential to a number of biological processes including gene expression and cell division, as well as to viral replication, tumorigenesis and tumor cell proliferation. The mechanism of nuclear transport has only recently been characterized in detail and has been shown to involve a number of discrete steps. Proteins that are destined to be transported into the nucleus contain within their amino acid sequence a short stretch of amino acids termed a nuclear localization sequence ("NLS"). These sequences may occur anywhere within the amino acid sequence and are typically four to about eight amino acids. These sequences are generally basic (i.e., positively charged) in nature, however, there has been no consensus sequence identified. Thus, there is a wide variety of these sequences that appear to be specific for particular proteins.

Within the cell, these NLSs may be either masked or unmasked by accessory proteins or by conformational changes within the NLS-containing protein. An NLS may be masked because it is buried in the core of the protein and not exposed on the surface of the protein. Unmasking of NLSs, and nuclear translocation of cytoplasmic proteins may be triggered by phosphorylation, dephosphorylation, proteolytic digestion, subunit association or dissociation of an inhibitory subunit, or the like. Accordingly, the masking and unmasking of NLSs provides a mechanism by which the transport of these cytoplasmic proteins into the nucleus may be regulated. For example, the transcription factor NF-AT contains nuclear localization sequences which allow NF-AT to translocate to the nucleus in the presence of intracellular calcium, but which are shielded by forming intramolecular associations with other domains in the NF-AT polypeptide in the absence of calcium.

Lee, H. C. and Bernstein, H. D. Proc. Natl. Acad. Sci. U. S. A. (2001), 98(6), 3471-3476 studied the mechanism involved for presecretory proteins such as maltose binding

protein (MBP) and outer membrane protein A (OmpA) that are targeted to the E. coli inner membrane by the molecular chaperone SecB, in contrast to the targeting of integral membrane proteins by the signal recognition particle (SRP). The authors found that replacement of the MBP or OmpA signal peptide with the first transmembrane segment of AcrB abolished the dependence on SecB for transport and rerouted both proteins into the SRP targeting pathway.

Some proteins contain cytoplasmic localization sequences (CLS), or nuclear export sequences, which ensure the protein remains predominantly in the cytoplasm. For example, Hamilton, M. H. et al., J. Biol. Chem. (2001), 276(28), 26324-26331 demonstrate that the ubiquitin-protein ligase (E3), hRPF1/Nedd4, a component of the ubiquitin-proteasome pathway responsible for substrate recognition and specificity, is capable of entering the nucleus, but the presence of a functional Rev-like nuclear export sequence in hRPF1/Nedd4 ensures a predominant cytoplasmic localization. The cytoplasmic domains of many membrane proteins contain sorting signals that mediate their endocytosis from the plasma membrane.

Heineman, T. C. and Hall, S. L. Virology (2001), 285(1), 42-49 studied three consensus internalization motifs within the cytoplasmic domain of VZV gB and determined that internalization of VZV gB, and its subsequent localization to the Golgi, is mediated by two tyrosine-based sequence motifs in its cytoplasmic domain. In mammalian cells and yeasts, amino acid motifs in the cytoplasmic tails of transmembrane proteins play a prominent role in protein targeting in the early secretory pathway by mediating localization to or rapid export from the endoplasmic reticulum (ER). Hoppe, H. C. and Joiner, K. A. Cell. Microbiol. (2000), 2(6), 569-578.

The mammalian endopeptidase, furin, is predominantly localized to the trans-Golgi network (TGN) at steady state. The localization of furin to this compartment seems to be the result of a dynamic process in which the protein undergoes cycling between the TGN and the plasma membrane. Both TGN localization and internalization from the plasma membrane are mediated by targeting information contained within the cytoplasmic domain of furin.

Voorhees, P., et al., EMBO J. (1995), 14(20), 4961-75 report that there are at least two cytoplasmic determinants that contribute to the steady-state localization and trafficking of furin. The first determinant corresponds to a canonical tyrosine-based motif, YKGL (residues 758-761), that functions mainly as an internalization signal. The second determinant consists of a strongly hydrophilic sequence (residues 766-783) that contains a large cluster of acidic residues (E and D) and is devoid of any tyrosine-based or di-leucine-based motifs. This

second determinant is capable of conferring localization to the TGN as well as mediating internalization from the plasma membrane.

The trans-Golgi network (TGN) plays a central role in protein sorting/targeting and the sequence SXYQRL can by itself confer significant TGN localization. Wong, S.H., and Hong, W. J. Biol. Chem. (1993), 268(30), 22853-62 report detailed mutagenesis of the 32-residue sequence of TGN38, an integral membrane protein confined mainly to the TGN, and determined that the Ser, Tyr, and Leu residues at positions 23, 25, and 28, respectively, are essential for TGN localization. When the cytoplasmic 32-residue sequence of TGN38 was fused to the ecto- and transmembrane domains of glycoporphin A (a surface protein), the resulting chimeric protein was localized to the TGN.

It is well recognized that certain proteins are either only active in a specific organelle, or are capable of different functions depending on their localization. For example, appropriate subcellular localization is crucial for regulation of NF- κ B function. Huang, T. T., et al., Proc. Natl. Acad. Sci. U. S. A. (2000), 97(3), 1014-1019, show that latent NF- κ B complexes can enter and exit the nucleus in preinduction states and identified a previously uncharacterized nuclear export sequence in residues 45-54 of I κ B α that was required for cytoplasmic localization of inactive complexes. It appears that NF- κ B/I κ B α complexes shuttle between the cytoplasm and nucleus by a nuclear localization signal-dependent nuclear import and a CRM1-dependent nuclear export and that the dominant nuclear export over nuclear import contributes to the largely cytoplasmic localization of the inactive complexes to achieve efficient NF- κ B activation by extracellular signals.

Nuclear import of classical nuclear localization sequence-containing proteins involves the assembly of an import complex at the cytoplasmic face of the nuclear pore complex (NPC) followed by movement of this complex through the NPC and release of the import substrate into the nuclear interior. In combination with Ran, two other soluble factors are thought to be absolutely required to mediate the nuclear import of a protein containing a classical or basic NLS into the nucleus. The first is karyopherin/importin α (Kap α), which binds a classical NLS and then forms a complex with karyopherin/importin β 1 (Kap β 1). Adam, S. A., and Gerace, L. (1991) *Cell* **66**, 837-847; Görlich, D., et al. (1994) *Cell* **79**, 767-778; Moroianu, J., et al. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2008-2011; Radu, A., et al. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1769-1773; Görlich, D., et al. (1995) *Curr. Biol.* **5**, 383-392; Chi, N. C., et al. (1995) *J. Cell Biol.* **130**, 265-274. Kap β 1 interacts with nuclear pore complex (NPC) proteins and appears to mediate movement of the import complex through the NPC via

these interactions. Rexach, M., and Blobel, G. (1995) *Cell* **83**, 683-692; Radu, A., Blobel, G., and Moore, M. S. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1769-1773; Iovine, M. K., Watkins, J. L., and Went, S. R. (1995) *J. Cell Biol.* **131**, 1699-1713; Radu, A., Moore, M. S., and Blobel, G. (1995) *Cell* **81**, 215-222. Another protein, p10/NTF2, has also been implicated in nuclear import, but its function may only be to take Ran into the nucleus, where it is subsequently needed to disassemble an incoming import complex. Moore, M. S., and Blobel, G. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10212-10216; Paschal, B. M., and Gerace, L. (1995) *J. Cell Biol.* **129**, 925-937; Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M., and Görlich, D. (1998) *EMBO J.* **17**, 6587-6598; Smith, A., Brownawell, A., and Macara, I. G. (1998) *Curr. Biol.* **8**, 1403-1406.

Although there is only one Kap α homologue in yeast (SRP1 or Kap60), vertebrate cells contain a number of proteins that can bind a classical NLS and share sequence homology (see Ref. Nachury, M. V., Ryder, U. W., Lamond, A. I., and Weis, K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 582-587, and references therein). These proteins have been given a variety of names but can be grouped into three major families. The Kap α 1 family contains the human protein NPI-1/importin α 1/karyopherin α 1/Rch2/hSRP1 and a second related protein importin α 6, in addition to the mouse S2 protein. Moroianu, J., et al., (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2008-2011; Cortes, P., et al., (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7633-7637; O'Neill, R. E., et al., (1995) *J. Biol. Chem.* **270**, 22701-22704; Kohler, M., et al., (1997) *FEBS Lett.* **417**, 104-108; Tsuji, L., et al., (1997) *FEBS Lett.* **416**, 30-34. The second family, Kap α 2, contains human Rch1/hSRP1/importin α 2/karyopherin α 2 and the mouse protein pendulin/PTAC 58. Görlich, D., Prehn, S., Laskey, R. A., and Hartmann, E. (1994) *Cell* **79**, 767-778; Cuomo, C. A., Kirch, S. A., Gyuris, J., Brent, R., and Oettinger, M. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6156-6160; Kussel, P., and Frasch, M. (1995) *Mol. Gen. Genet.* **248**, 351-363; Imamoto, N., Shimamoto, T., Takao, T., Tachibana, T., Kose, S., Matsubae, M., Sekimoto, T., Shimonishi, Y., and Yoneda, Y. (1995) *EMBO J.* **14**, 3617-3626; K., Mattaj, J. W., and Lamond, A. I. (1995) *Science* **268**, 1049-53. The third family, Kap α 3, consists of the two human proteins, QIP-1/importin α 3 and KPNA3/hSPR1 γ /hSRP4, and the mouse proteins Q1 and Q2. Nachury, M. V., et al., (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 582-587; Kohler, M., et al., (1997) *FEBS Lett.* **417**, 104-108; Tsuji, L., et al., (1997) *FEBS Lett.* **416**, 30-34; Takeda, S., et al., (1997) *Cytogenet. Cell Genet.* **76**, 87-93; Seki, T., et al., (1997) *Biochem. Biophys. Res. Commun.* **234**, 48-53; Miyamoto, Y., et al., (1997) *J. Biol. Chem.* **272**, 26375-26381. Each of these classes share about 50% homology with each other

and to the yeast SRP1, and each of these mammalian proteins has been shown to be capable of mediating the import of one or more classical NLS-containing proteins. Nachury, M. V., et al., (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 582-587; Sekimoto, T., et al., (1997) *EMBO J.* **16**, 7067-7077; Nadler, S. G., et al., (1997) *J. Biol. Chem.* **272**, 4310-4315; Prieve, M. G., et al., (1998) *Mol. Cell. Biol.* **18**, 4819-4832.

Stat-1 import is mediated by Kap α 1/NPI-1 but not Kap α 2/Rch1, but activated Stat-1 appears to bind to a COOH-terminal region of Kap α 1 distinct from the NLS binding Armadillo repeats. The binding differences of the different Kap α s to RCC1 observed appear to be due solely to the NLS on RCC1 and therefore probably due to the NLS binding region of Kap α 3. Sekimoto, T., et al., (1997) *EMBO J.* **16**, 7067-7077. Kamei, Y., et al., (1999) *J. Histochem. Cytochem.* **47**, 363-372 showed that, in mice, the Kap α 3 homologue is expressed in many tissues and theorized that Kap α 3 may play a role in importing "a limited number of unique karyophilic proteins, such as helicase Q1." The results provided by Talcott, B. and Moore, M.S., 2000 *J Biol Chem*, 275(14) 10099-10104 suggest that RCC1 should be included in the group of proteins that use Kap α 3 to mediate their nuclear import.

USP 6,191,269 teaches the existence of a nuclear localization sequence contained within the cDNA sequence of the N-terminal IL-1 alpha propiece, T76-NGKVLKKRRL, which had characteristics of a nuclear localization sequence (NLS) and could mediate nuclear localization of the propiece (Stevenson et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:508-13). Introduction of the cDNA encoding the N-terminal IL-1.alpha. propiece into cultured mesangial cells resulted in nuclear accumulation (Stevenson et al. id).

USP 5,877,282 teaches that the antennapedia homeodomain signal sequence peptide is the amino acid sequence RQIKIWFQNRRMKWKK; the fibroblast growth factor signal sequence peptide is AAVALLPAVLLALLA; the HIV Tat signal sequence peptide is the amino acid sequence CFITKALGISYGRKKRRQRRRPQGSQTH.

Schwartz, S.R., et al., *Science* 285:1569-1572 (1999) report delivery of an ip injected reporter protein, 116 kD beta-galactosidase, as a TAT fusion protein into tissues and across the blood-brain barrier. Schwartze used an 11 amino acid protein transduction domain (PTD) derived from HIV tat protein with an N-terminal fluorescein isothiocyanate (FITC)-Gly-Gly-Gly-Gly motif. The authors report that earlier attempts to transduce beta-Gal chemically cross-linked to the TAT PTD resulted in sporadic and weak beta-Gal activity in a limited number of tissues. They speculate that the improved transduction was due to the in-frame fusion and purification strategy used.

Nuclear localization of IFN γ is mediated by a polybasic NLS in its C terminus, which is required for the full expression of biological activity of IFN γ , both extracellularly and intracellularly. Subramaniam, Prem S., et al., *J. Cell Sci.* (2000), 113(15), 2771-2781. This NLS is thought to play an integral intracellular role in the nuclear translocation of the transcription factor STAT1 α activated by IFN γ because treatment of IFN γ with antibodies to the C-terminal region (95-133) containing the NLS blocked the induction of STAT1 α nuclear translocation, but these antibodies had no effect on nuclear translocation of STAT1 α in IFN α treated cells. A deletion mutant of human IFN γ , IFN γ (1-123), which is devoid of the C-terminal NLS region was biologically inactive, but was still able to bind to the IFN γ receptor complex on cells with a K_d similar to that of the wild-type protein. Deletion of the NLS specifically abolished the ability of IFN γ (1-123) to initiate the nuclear translocation of STAT1 α , which is required for the biological activities of IFN γ following binding to the IFN γ receptor complex. A C-terminal peptide of murine IFN γ , IFN γ (95-133), that contains the NLS motif, induced nuclear translocation of STAT1 α when taken up intracellularly by a murine macrophage cell line. Deletion of the NLS motif specifically abrogated the ability of this intracellular peptide to cause STAT1 α nuclear translocation. In cells activated with IFN γ , IFN γ was found to as part of a complex that contained STAT1 α and the importin- α analog Npi-1, which mediates STAT1 α nuclear import. The tyrosine phosphorylation of STAT1 α , the formation of the complex IFN γ /Npi-1/STAT1 α complex and the subsequent nuclear translocation of STAT1 α were all dependent on the presence of the IFN γ NLS.

The peptide representing amino acids 95-132 of IFN- γ (IFN- γ (95-132)), containing the polybasic sequence ¹²⁶RKRKRSR¹³², was capable of specifying nuclear uptake of the autofluorescent protein, APC, in an energy-dependent fashion that required both ATP and GTP. Nuclear import was abolished when the above polybasic sequence was deleted.

Subramaniam, P., et al., 1999 *J Biol Chem* 274(1) 403-407. A peptide containing the prototypical polybasic NLS sequence of the SV40 large T-antigen was also able to inhibit the nuclear import mediated by IFN- γ (95-132), suggesting that the NLS in IFN- γ may function through the components of the Ran/importin pathway utilized by the SV40 T-NLS. Intact IFN- γ , when coupled to APC, was also able to mediate its nuclear import, and this nuclear import was blocked by the peptide IFN- γ (95-132) and the SV40 T-NLS peptide, suggesting that intact IFN- γ was also transported into the nucleus through the Ran/importin pathway.

Nuclear proteins are imported into the nucleus through aqueous channels that span the nuclear envelope called nuclear pore complexes (NPCs). Although ions and molecules less than ~20-40 Da can diffuse passively through the nuclear pore complexes, larger proteins are transported by saturable pathways that are energy- and signal-dependent. The signals that specify nuclear protein import (NLSs) are commonly short stretches of amino acids rich in basic amino acid residues, although other classes of NLSs have been described recently. The initial step in the import of proteins containing basic amino acid-type NLSs occurs in the cytosol, where the NLS-containing proteins are bound to a receptor (variously called the NLS receptor, importin α , and karyopherin (13). The substrate-receptor complex then associates with the cytoplasmic face of the nuclear pore complexes, and with the participation of other cytosolic factors, is transported through a gated channel in the nuclear pore complexes to the nuclear interior. The *in vivo* events of NLS-mediated nuclear import can be duplicated in an *in vitro* system using digitonin-permeabilized cells supplemented with cytosolic extracts and ATP (14). Transport in this *in vitro* assay is blocked by the same inhibitors that block *in vivo* import, is rapid, and is easily quantified.

The NLS sequence NYKKPKL in the N-terminus of fibroblast growth factor (FGF)-1, the precursor for acidic FGF, has been proposed to affect the long term activities of FGF-1 through its function as a nuclear translocation signal or its role in stabilization of the structure required to sustain binding and activation of the transmembrane receptor kinase. Luo, Y., et al., *J. Biol. Chem.* (1996), 271(43), 26876-26883. For example, concurrent with a marked increase in dependence on exogenous heparin for optimal activity, sequential deletion of residues in the NYKKPKL sequence in FGF-1 resulted in a progressive loss of thermal stability, resistance to protease, mitogenic activity, and affinity for the transmembrane receptor. The largest change resulted from deletion of the entire sequence through the lysine-leucine residues. In the presence of sufficiently high concentrations of heparin, the deletion mutants exhibited mitogenic activity equal to wild-type FGF-1.

Although FGF-1 contains an NTS, nuclear translocation requires an exogenous and not an endogenous pathway. The NTS of FGF-1, NYKKPKL, is able to direct the expression of the bacterial β -galactosidase (β gal) gene to the nucleus of transfected NIH 3T3 cells, but this NTS is unable to target either FGF-1 itself or a FGF-1- β gal fusion protein into the nucleus, suggesting that FGF-1 may contain an additional sequence which prevents endogenously expressed FGF-1 from being translocated into the nucleus. Zhan, X., et al., *Biochem. Biophys. Res. Commun.* (1992), 188(3), 982-91.

Interferon- γ (IFN- γ), a protein that uses the Jak-Stat pathway for signal transduction, translocates rapidly to the nucleus in cells treated extracellularly with the cytokine. An NLS has been identified and characterized in the C-terminus of human and murine IFN- γ . Larkin, J., et al., J. Interferon Cytokine Res. (2001), 21(6), 341-348 report that human IFN- γ (HuIFN- γ) contains a second NLS at an upstream site. The primary sequence, analogous with the NLS sequence identified in murine IFN- γ , representing amino acids 122-132 of HuIFN- γ was capable of mediating the nuclear import of the autofluorescent protein allophycocyanin (APC) in an energy-dependent manner. The second sequence, representing amino acids 78-92 of HuIFN- γ , was also capable of mediating the nuclear import of APC in an energy-dependent manner but to a greatly reduced extent. The nuclear import of both sequences conjugated to APC was strongly blocked by competition with unconjugated HuIFN- γ (122-132). Competition by the sequence HuIFN- γ (78-92) effectively blocked the import of APC-conjugated HuIFN- γ (78-92) but, at the same concentration, was not capable of inhibiting the nuclear import of APC-conjugated HuIFN- γ (122-132), suggesting that HuIFN- γ (78-92) was a less efficient NLS than HuIFN- γ (122-132). This is consistent with >90% loss of antiviral activity of HuIFN- γ lacking the downstream NLS in 122-132. The nuclear import of APC-conjugated HuIFN- γ (122-132) was inhibited by a peptide containing the prototypical polybasic NLS of the SV40 T NLS, which suggests that the same Ran/importin cellular machinery is used in both cases.

There appears to be strong conservation of the NLS motif as a mechanism for nuclear localization. Evolution seemed to have used part of the existing DNA-binding mechanism when compartmentalizing DNA-binding proteins into the nucleus. Cokol, M., et al., EMBO Rep. (2000), 1(5), 411-415 estimate that greater than 17% of all eukaryotic proteins may be imported into the nucleus, and after analyzing a set of 91 experimentally verified NLSs from the literature and expanding this set to 214 potential NLSs through iterated "in silico mutagenesis". This final set matched in 43% of all known nuclear proteins and in no known non-nuclear protein. Cokol et al found an overlap between the NLS and DNA-binding region for 90% of the proteins for which both the NLS and DNA-binding regions were known, but only 56 of the 214 NLS motifs overlapped with DNA-binding regions. These 56 NLSs enabled a de novo prediction of partial DNA-binding regions for approximately 800 proteins in human, fly, worm and yeast.

More recently, it has been reported that NLS signal peptide can induce structural changes of DNA. The plant enzyme, glutaminyl-tRNA synthetase (GlnRS) from *Lupinus*

luteus, contains an NLS at the N-terminal, a lysine rich polypeptide, KPKKKKKEK. Krzyzaniak, A., et al., Mol. Biol. Rep. (2000), 27(1), 51-54. Two synthetic peptides (20 and 8 amino acids long), derived from the NLS sequence of lupin GlnRS interact with DNA. In addition, the shorter 8 amino acid peptide caused the DNA to change its conformation from the B to the Z form. This observation clearly suggests that the presence of the NLS polypeptide in a leader sequence of GlnRS is required not only for protein transport into nucleus but also for regulation of a gene expression. This is the first report suggesting a role of the NLS signal peptide in structural changes of DNA.

Typically there is strong conservation of the NLS sequence within species. For example, the NLS in the N-terminal region of Smad 3 protein, the major Smad protein involved in TGF- β signal transduction, has a basic motif Lys⁴⁰-Lys-Leu-Lys-Lys⁴⁴, which is conserved among all the pathway-specific Smad proteins, and is required for Smad 3 nuclear import in response to ligand. Smad proteins are intracellular mediators of transforming growth factor- β (TGF- β) and related cytokines. Xiao, Z., et al., J. Biol. Chem. (2000), 275(31), 23425-23428 identified the role the NLS plays in nuclear localization. The authors demonstrated that the isolated Smad 3 MH1 domain displays significant specific binding to importin β , which is diminished or eliminated by mutations in the NLS. Full-size Smad 3 exhibits weak but specific binding to importin β , which is enhanced after phosphorylation by the type I TGF- β receptor. In contrast, no interaction was observed between importin α and Smad 3 or its MH1 domain, indicating that nuclear translocation of Smad proteins may occur through direct binding to importin β . The authors conclude that activation of all of the pathway-specific Smad proteins (Smads 1, 2, 3, 5, 8, and 9) exposes the conserved NLS motif, which then binds directly to importin β and triggers nuclear translocation.

In all cells, the lipid bilayer of cell membranes serves as a selective barrier for the passage of charged molecules, with the internalization of hydrophilic macromolecules being achieved through classical transport pathways (Hawiger, J., *Curr Opin Chem Biol.* 3, 89-94 (1999), Schwarze, S.R., et al., *Trends in Cell Biology* 10, 290-295 (2000)). These classical mechanisms of internalization involve receptor-mediated endocytosis or transporter dependent uptake (Cleves, A.E., *Current Biology* 7, R318-R320 (1997)). In contrast, an increasing number of molecules have been discovered that lack classical import and/or export signals (Cleves, A.E., *Current Biology* 7, R318-R320 (1997)). These molecules gain direct access to either cytoplasmic or nuclear compartments using unconventional processes of which the mechanisms remain largely unknown. These novel mechanisms are generally termed

“nonclassical” and refer to transport pathways being used that are atypical. Relevant examples of this latter type are found in the gene-encoded proteins of HIV-1 TAT (Frankel, A.D. and Pabo, C.O. *Cell* **55**, 1189-1193 (1988)), herpes virus VP22 (Elliott, G. and O'Hare, P. *Cell* **88**, 223-233 (1997)), and Antennapedia, Antp (Derossi, D., et al., *J. Biol. Chem.* **269**, 10444-10450 (1994)). It is now well established that the full-length proteins of HIV-1 TAT (Helland D.E., et al., *J Virol* **65**, 4547-4549 (1991)), and VP22 (Pomeranz L.E. and Blaho J.A., *J Virol* **73**, 6769-6781 (1999)) rapidly translocate into and out of cellular membranes. In fact, distinct peptide regions have been identified within both of these proteins that are capable of translocating into cellular compartments either alone or in combination with chimeric cargo peptides, and proteins (Lindgren, M., et al., *Trends Pharmacol Sci.* **3**, 99-103 (2000), Derossi, D., et al, *Trends Cell Biol.*, **8**, 84-87 (1998), Prochiantz A., *Current Opinion in Cell Biology* **12**, 400-406 (2000), Steven R. Schwarze, S.R., et al., *Trends in Cell Biology* **10**, 290-295 (2000)). In contrast, full-length Antp protein has not been shown to traverse biological membranes; however, a 16 amino acid synthetic peptide derived from within its coding region does possess potent membrane penetrating abilities (Derossi, D., et al, *Trends Cell Biol.*, **8**, 84-87 (1998)). The accepted view of atypical transport used by these molecules has been termed “transduction” (Schwarze, S.R., et al., *Trends in Cell Biology* **10**, 290-295 (2000)), and is currently defined as an extremely rapid membrane transport pathway that is receptor and energy independent, and can occur at 4 C in all cell types (Schwarze, S.R. and Dowdy, S.F. *Trends Pharmacol. Sci.* **21**, 45-48 (2000)). Interestingly, these three proteins are all nuclear proteins involved in transcriptional regulation, and their respective transducing peptides consist of strings of amino acids rich in arginine and lysine (Lindgren, M., et al., *Trends Pharmacol Sci.* **3**, 99-103 (2000), Schwarze, S.R. and Dowdy, S.F. *Trends Pharmacol. Sci.* **21**, 45-48 (2000)). However, irrespective of these similarities, these transducing peptides possess many different characteristics such as amino acid sequence, length of the sequence, cellular localization, and potency of membrane penetration. Thus, though each transducing sequence can penetrate cells and tissues, it has not been established whether they use the identical atypical transport mechanisms.

Finally, USP 6,022,950 teaches the use of a hybrid molecule of a portion of the binding domain of a cell-binding polypeptide ligand effective to cause said hybrid protein to bind to a cell of an animal, a translocation domain of naturally occurring protein which translocates said third part across the cytoplasmic membrane into the cytosol of the cell; and a chemical entity to be introduced into the cell. However, the patent teaches translocation domains of toxins. Naturally-occurring proteins which are known to have a translocation

domain include diphtheria toxin and *Pseudomonas* exotoxin A, and may include other toxins and non-toxin molecules, as well. The translocation domains of diphtheria toxin and *Pseudomonas* exotoxin A are well characterized (see, e.g., Hoch et al., Proc. Natl. Acad. Sci. USA 82:1692-1696, 1985; Colombatti et al., J. Biol. Chem. 261:3030-3035, 1986; and Deleers et al., FEBS 160:82-86, 1983), and the existence and location of such a domain in other molecules may be determined by methods such as those employed by Hwang et al., Cell 48:129-136, 1987; and Gray et al., Proc. Natl. Acad. Sci. USA 81:2645-2649, 1984.

Given the considerable body of literature teaching control mechanisms of cellular localization, the proteins involved in regulation of intracellular transport, the different properties and control mechanisms for plasma membrane and the nuclear envelope, it is unexpected that polypeptides derived from mammalian proteins could transduce through the plasma membrane using nonclassical mechanisms and thus could be useful as membrane penetrating peptides useful as *in vitro*, *ex vivo* and *in vivo* delivery devices of a compound of interest. There is also considerable literature teaching non-protein derived methods for delivering a compound of interest into cells, for example electroporation, membrane fusion with liposomes, high velocity bombardment with DNA-coated microprojectiles, incubation with calcium-phosphate-DNA precipitate, DEAE-dextran mediated transfection, infection with modified viral nucleic acids, and direct microinjection into single cells, usually ova and the like. Each of these methods is relatively inefficient, resulting in relatively low percentage of the cells containing the delivered compound of interest and most of the methods are clearly not capable of realistic *in vivo* delivery. Many of the methods are toxic to the cells, resulting in relatively high apoptosis. Therefore, there is a considerable need for simple and more efficient delivery of compounds of interest into cells.

SUMMARY OF THE INVENTION

The present invention is directed to polypeptides derived from mammalian and yeast proteins useful as a carrier for *in vitro*, *ex vivo* and *in vivo* delivery a compound of interest. The invention also provides compositions containing the same, and methods of delivering a compound of interest *in vitro*, *ex vivo* and *in vivo*.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. (A). Schematic diagram of hPER1 fusion constructs showing the locations of the PAS, cytoplasmic localization, and nuclear localization sequence (NLS, but indicated as nuclear localization domain (NLD) in Figure). The name and the position of the fusion

constructs are listed on the left. The number indicates the first and last amino acid residues in the hPER1 protein. The principal sites of accumulation of each fusion protein are summarized on the right, **(n)** nuclear, **(no)** nucleoli, **(c)** cytoplasmic, **(diff)** diffuse. All constructs were N-terminally tagged with EYFP. The alignment human and mouse PER1-NLS is shown at the bottom.

Figure 1. (B). Cellular localization of hPER1 fusion proteins as described in Figure 1A, above, in living cells. CHO cells were transiently transfected with the fusion constructs indicated on the top of each panel and the subcellular localization of EYFP reporters (green) was directly visualized using fluorescent microscopy 10 h post-transfection. EYFP vector alone is used as control (see 5. EYFP-VECTOR)

Figure 2. (A). Membrane penetration assay in CHO cells. N-terminal biotinylated synthetic peptides hPER1-PTD, Flag-hPER1-PTD, Flag-TAT-PTD (positive control), and Flag-Flag (negative control) were assayed for their ability to penetrate cellular membranes in living CHO cells in culture. The subcellular localization of internalized peptides was determined using a two color staining method, either Streptavidin-Alexa 594 (red) or anti-flag mAb (green). The third column is an overlay (yellow). Confocal microscopy was employed to further confirm intracellular and intranuclear localization. Single section of confocal imaging is shown.

Figure 2. (B). Nuclear targeting of biotinylated peptides hPER1-NLD (also known as hPER1-PTD) compared with TAT-PTD and Flag-Flag (negative control) using Streptavidin Alexa-594 fluorescence (green). Hoechst 33258 at 5ng/ml was used to stain the nucleus (blue, middle column). The third column is an overlay of confocal imaging.

Figure 3. Alanine scanning of hPER1-PDTs. Biotinylated hPER1-NPDs were synthesized with a single amino acid substitution at the indicated position with an alanine and assayed for membrane penetration in CHO cells. Cells were incubated for 10 minutes at 37 C at a peptide concentration of 10 μ M followed by washing, fixation, permeabilization, and then detected with labeled Streptavidin Alexa-594 (red, 2 μ g/ml) for 15 minutes at the RT. Control peptide was from hPER1 N-terminal amino acids residues 486-500.

Figure 4. Activation of serotonin 5HT2A receptor with hPER1-MPP fusion peptide. **(A).** hPER1-MPP and TAT-PTD peptides were synthesized alone or in fusion with either the first intracellular loop II (SLEKKLQNATN), or the C-terminal Transmembrane 7 domain, TM7 (KTYRSAFSRYIQYKKNKPLQLI) derived from the 5HT2A receptor, genebank

accession numbr, M86841). Receptor activities was assayed using standard FLIPR analysis and measuring endogenous and exogenous Ca^{+2} levels. Peptide designations are as follows: **T** (TAT-PTD), **P** (hPER1-MPP), **I1** (intracellular loop 1), **T-I1** (TAT-PTD-I1), **P-I1** (hPER1-MPP-I1), **TM7** (C-terminal domain), **TTM7** (TAT-PTD-TM7), **PTM7** (hPER1-MPP-TM7), and **S** (Serotonin).

Figure 4 (B). Dose response of PTM7 (closed circles) and TTM7 (closed diamonds) peptides. Serotonin (control, open triangle) was used at the maximum receptor stimulatory concentration of 10 μM .

Figure 5. Identification of additional PTDs. Putative PTD sequences were searched using a combined bioinformatics method that included SwissPro, PRF, PIR-Protein info Resource, PDB with peptides sequences translated from the annotated protein coding region in GenBank with "transcription factor" as the key word. We initially searched for all known or putative NLS's. Secondly, we employed the PHI-BLAST (Pattern-Hit Initiate BLAST) to search for the degenerative pattern occurrence $[\text{R}/\text{H}/\text{K}]\text{-}[\text{R}/\text{H}/\text{K}]\text{-}[\text{R}/\text{H}/\text{K}]\text{-}[\text{R}/\text{H}/\text{K}]$, $(\text{X})_n$ where n is an integer of 4 or larger and X each time is independently selected to be either arginine, histidine, or lysine. 7374 putative PTD sequences were identified. From the two searches we synthesized (A) biotinylated peptides to these sequences or (B) created in frame fusion proteins with GFP and transfected CHO cells. 9 of the 12 peptides were found to transduce, and all sequences localize to the nucleus in transfected cells. hPER1-PTD, hPER3-PTD, and TAT-PTD peptides were used as positive controls. Six positive sequences and 2 negative sequences are shown. Numbers represent the amino acid residues within the parental protein sequence and Gene bank accession numbers for these proteins are indicated as follows: (M24899, human Thyroid hormone apha-1; L12699, human Homeobox protein Engrailed 1 HME1; X16416, human Proto-oncogene tyrosine protein kinase ABL1;; Q02575, human HEN1/NSLC1; Q02577, human HEN2/NSLC2; AAA74561, rat HNF-3; CAB65887, *Drosophila* cAMP dependent transcription factor). Three negative peptides are (V01512, c-Fos; AAD53184, human cyclin L ania-6a; CAB66914, *Arabidopsis* β -zip transcription factor).

Figure 6. hPER-PTD cargo's β -Galactosidase into cells: At least one feature of HIV TAT transducing peptide is its ability to cargo proteins into cells and tissues. We therefore sought to determine if hPER1 transducing peptide could cargo beta galactosidase into cells. To perform this experiment, we followed a protocol by Frankel *et al.* 1989 (19):7397-401, whereby, we chemically linked hPER1-PTD or hPER-PTD R7A to full length β -galactosidase and assayed for the ability of these conjugates and beta-galactosidase protein

alone to transduce into CHO cells. As shown in the figure 6, left, cells incubated with hPER-PTD β -galactosidase fusion showed positive enzymatic activity for β -galactosidase as indicated by the blue color in the cells after the addition of X-gal. However, neither hPER-MPP R7A β -galactosidase (center) nor β -galactosidase protein (right) alone was able to enter the cells as indicated by a no blue staining reactivity after the addition of X-gal. These data indicate that like TAT peptide, hPER1-PTD can cargo a large (120 kD) protein into cells.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on discovery that human Period1 (hPER1) protein contains an NLS which has now also been identified as an MPP and is useful as a delivery device for intracellular delivery of a compound of interest. hPER1 is involved in regulation of the circadian rhythm and the capacity of hPER1 to translocate to adjacent cells may be critical to its overall biological function of regulating circadian rhythm. The NLS identified within hPER1 does not fit within previously identified NLS sequences, and its identification has resulted in identification of an algorithm for searching for other NLS sequences which may also function as MPPs.

Period 1 (hPER1) is a nuclear protein involved with transcriptional regulation. It is an essential component in the “gears” of the biological clock (Brown, S.A., and Schibler, U., *Current Opinon in Genetics & Development* **9**, 588-594 (1999), Dunlap, J.C., *Cell* **96**, 271-290 (1999)), and studies in mice have shown that nuclear entry of PER1 is essential for the down regulation of CLOCK/BMAL transcriptional complexes (Gekakis N, et al., *Science* **280**, 1564–1569. (1998), Yagita, K., et al., *Genes Dev* **14**,1353-1363 (2000), Lowrey, P.L., et al., *Science* **288**, 483-492 (2000)). However, to date, the functional NLS for human PER1 has not been elucidated. The present inventors identified the NLS within hPER1, and demonstrate that the **16 amino acid** and 13 amino acid sequence, see Figure 3.hPER1-NLS peptide, hPER1-MPP, has potent membrane penetrating ability. This work results in the identification of four additional MPPs also derived from nuclear proteins.

PER1 is a central component in the circadian clock, and its nuclear entry plays an important role in the regulation of daily oscillations (Jin, X., et al., *Cell* **96**, 57-68 (1999), Sangoram, A.M., et al., *Neuron* **21**, 1101-13 (1998)). Using deletion and fusion protein analysis, we identified a NLS that is necessary and sufficient for hPER1 nuclear localization. This functional analysis was necessary because the NLS of hPER1 does not conform to classical nuclear localizing consensus motifs; and therefore, was not identified using standard

NLS search procedures. We show that a single copy of hPER1-NLS is sufficient for inducing nuclear localization of a reporter protein and of tagged hPER1 fragments (P1-F2 to P1-F7) in transfected cells. The PER1-NLS is located between amino acids (830-845) of hPER1, is embedded within a string of 13 amino acids rich in arginine, histidine, and lysine (see Table 1) that is not found in other PERs or other nuclear proteins in available databases. Therefore, though PERs 2 and 3 are nuclear proteins (Jin, X., et al., *Cell* 96, 57-68 (1999)), they apparently use alternative sequences and or mechanisms for their nuclear import.

Peptide fragments of a limited number of nuclear proteins that are rich in basic residues have been shown to penetrate into cellular membranes in a receptorless, energy-independent fashion. Sequences from three such proteins, TAT, Antp, and VP22 have been demonstrated to possess the ability to penetrate and cargo fusion molecules into cells and tissues by an as yet undefined mechanism. See, for example, USP 5,804,604, 5,747,641, 5,674,980, 5,670,617 and 5,652,122 issued to Frankel et al., which teach the use of a nine-amino acid HIV TAT-derived polypeptide (Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg) for intracellular delivery of cargo molecules.

The similarities between hPER1, the hPER1-NLS, and other MPPs prompted us to investigate whether or not hPER1-MPP could have membrane penetrating capability. The immunohistochemical and cytological data presented herein indicates that the hPER1-MPP functions as a MPP in a variety of cell types. hPER1-MPP demonstrated intense focal staining in the nuclear plasma as well as in the nucleolus, suggesting that the subnuclear address of hPER1-MPP is different from the hPER1 (P1-FL) protein that was diffused in the nucleus but not concentrated in the nucleolus. The cellular penetration of hPER1-MPPs is not blocked even under the conditions of reversing the sequence (reversed hPER1-MPP), adding negatively charged residues or pre-fixing cells with 4% PFA, unpublished observation, the latter supports the idea that penetration is receptor and membrane independent. These results are in contrast to other peptide classes that have been described that are derived from signal peptide sequences (Hawiger, J., *Curr Opin Immunol.* 9, 189-94 (1997)), DNA antibodies (Deng, S.X., et al., *Int Immunol.* 12, 415-423 (2000)), and other protein domains (Lindgren, M., et al., *Trends Pharmacol Sci.* 3, 99-103 (2000)) that bind and cross the cell membranes using slow, temperature, energy, and receptor dependent mechanisms.

The identification of other MPPs, has been limited by our lack of understanding the mechanisms and structural requirements necessary for membrane peptide penetration. The likelihood that a specific peptide structure and/or charge is important for membrane penetration is demonstrated in the alanine scanning experiments whereby a single amino acid

change at arginine 7 appears to be critical for MPP potential. By comparing wild-type hPER-MPP to modified P1- R7A, in live cells or pre-fixed and permeabilized cells (data not show), P1-R7A is only defective in penetration but not in nuclear targeting once the cells have been permeabilized. This finding suggests that arginine 7 has a major role in structure based penetration, and thus provides a useful model for the future structure-function studies. No structural determinants for TAT peptide have been described, but in the case of Antp, replacing the two tryptophan residues with two phenylalanines abolishes penetration (Le Roux, I., et al., *Proc Natl Acad Sci U S A.* 90, 9120-9124 (1993)). Since hPER1-MPP does not contain any tryptophan residues, membrane penetration between these two peptides may occur by different mechanisms.

Full-length HIV TAT and VP22, both of which lack classical secretory signal sequences and are therefore exported by non-classical mechanisms, can also be imported “by transduction”, into cells in a non-classical manner (Prochiantz A., *Current Opinion in Cell Biology* 12, 400-406 (2000)). Therefore, it is interesting to speculate that perhaps hPER1 distributes circadian clock information to adjacent SCN neurons or to circadian output pathways by “transduction” mechanisms similar to full-length TAT and VP22 proteins. However, simply having membrane penetrating sequences within the body of a protein does not necessarily confer membrane penetrating capability, as full-length Antp protein is neither exported from nor imported into cells. Thus, the non-classical penetration of the Antp peptides into the cells is unlikely to have physiological relevance, and like Antp, there is no evidence to suggest that full-length hPER1 is a cell membrane penetrating protein. However, these findings did encourage us to search for other MPP-containing proteins. By searching protein databases with an algorithm designed to identify strings of basic residues within nuclear proteins, we uncovered hundreds of proteins that contained potential membrane penetrating peptide regions and found 4 additional MPPs from several species (see Fig. 5). These and additional database mining searches suggest that MPP-like sequences are common, and present within a wide variety of proteins. However, like many putative NLSs that do not always confer nuclear localization when fused to reporter sequences (Moroianu, J., *J Cell Biochem.* 32-33, 76-83 (1999)), any potential MPPs must be functionally determined experimentally. Though it seems clear that either transducing or non-transducing proteins can encode MPP regions, the interesting question that remains is whether or not proteins containing MPP-like sequences use these domains to rapidly translocate intracellularly into cellular domains to activate normal physiological processes. The efficiency associated with the transduction phenomena might be particularly useful where the rapid delivery of

intercellular information is critical, as may be the case in cell synchronization, development, and differentiation paradigms.

The ability for MPPs to cargo molecules to intracellular compartments is becoming well-established (Lindgren, M., et al., *Trends Pharmacol Sci.* 3, 99-103 (2000), Derossi, D., et al, *Trends Cell Biol.*, 8, 84-87 (1998)). Similar to other MPPs, hPER1-MPP and other MPPs identified herein can deliver compounds of interest, such as large molecules, i.e., peptides and proteins, lipids, polysaccharides, other organic molecules, rapidly and efficiently into cells. The data presented herein demonstrates that hPER1-MPP in fusion with either serotonergic and/or adrenergic 7TM-receptor derived peptides mimic the effects of ligand activated receptors (see Fig. 4, and data not shown), confirming that hPER1-MPP translocates compounds of interest to intracellular compartments, and supports the idea that physiologically relevant signaling can be initiated by MPPs linked to compounds of interest. Using the methods described herein, the present invention may be expanded to provide target validation using MPPs linked to targets, and/or therapeutic strategies using MPPs linked to specific enzymes or receptors as a method of altering, correcting or compensating for dysfunctional enzyme performance or within pathways. In addition, therapeutic strategies using MPPs linked to specific receptors may be used as a method of altering, correcting or compensating for dysfunctional receptor, low expression of normal or abnormal receptors.

Taken together, the results provided herein demonstrate an MPP encoded by a mammalian protein and more specifically, a human nuclear protein, whose cellular penetration is membrane independent and likely depends on the peptide structure. hPER1-MPP targets to specific subnuclear sites, but has the potential to efficiently deliver other macromolecules to intracellular targets.

More importantly, this invention also provides the first example for mapping a novel MPP based on a NLS domain, and suggests that many MPP-like regions are contained within a wide variety of proteins. The data provided herein demonstrate that an MPP may be based on part of an NLS, or overlap with part of the NLS, or alternatively, may be a novel peptide.

Methods of identifying NLS sequences are well known in the art, and include NLSs previously identified as conferring the ability of the native protein to enter the nucleus, or is a putative NLS based on substantial sequence homology with a previously identified NLS. Alternatively, the NLS may be identified by sequence deletion experiments. See for example, Luo JC, Shibuya M A variant of nuclear localization signal of bipartite-type is required for the nuclear translocation of hypoxia inducible factors (1alpha, 2alpha and 3alpha). *Oncogene*.

2001 Mar 22;20(12):1435-44 or Hodel MR, Corbett AH, Hodel AE. Dissection of a nuclear localization signal. J Biol Chem. 2001 Jan 12;276(2):1317-25.

Preferred membrane penetrating peptides (MPPs, also known as peptide transduction domain or 'PTD') of the present invention are small polypeptides, and may be derived from an NLS, or overlapping with an NLS, of a mammalian or yeast protein. Preferred mammalian proteins are those of human, primate, murine or rat species. It is generally preferred to use the same species for the NLS-derived protein as the cell to be treated. Human species are especially preferred as the NLS-derived protein when being used to treat human cells. NLSs may be found within a broad class of enzymes, and is not limited to nuclear proteins, transcription factors, cytokines and kinases. Preferred MPPs are those derived from nuclear proteins or transcription factors. Alternatively, MPPs of the present invention are small polypeptides comprising a sequence $-(X-X-X-X)_n-$ where n is an integer 1 to 7, and X each time is independently selected from the group consisting of arginine, histidine or lysine. It is preferred that small MPPs are used, and therefore, it is preferred that n is an integer 1 to 5, and more preferred that n is an integer 1 to 3. Selected embodiments of suitable MPPs are provided in Table 1 and Example 5.

The MPP and/or compound of interest may be chemically synthesized separately, for example, by chemical synthetic routes and using commercially available reagents. Alternatively, if the MPP and/or compound of interest is a polypeptide, it may be synthesized by recombinant technology and purified according to known methods. Host cells, cloning vectors, promoters and oligonucleotide linkers are well known and commercially available. Methodologies for using recombinant technology and purification methods are also well known, see Current Protocols in Molecular Biology, 4 Vols. Wiley. Generally, recombinant technology is preferred, as it is more amenable to larger scale production and is more economical for mass production. MMPs may be obtained by specific protease degradation of a precursor proteins.

The compound of interest may be attached or linked to the MPP via chemical crosslinking at the N- or C-terminus of the MPP to create a conjugated (also referred to a fusion) MPP and compound of interest, for example, via disulfide or ester linkages. In an alternative embodiment, if the compound of interest is a peptide, the peptide may be synthesized by recombinant technology with a host cell with an expression vector encoding a fusion of the MPP sequence and the compound of interest under conditions to permit expression of the vector and obtaining the fusion MPP and compound of interest.

In another embodiment, the MPP and the compound of interest may be attached or linked via a chemical linker. Chemical linkers are well known in the art, and include but are not limited to dicyclohexylcarbodiimide (DCC), N-hydroxysuccinimide (NHS), maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), N-ethyloxycarbonyl-2-ethoxy-1,2-dihydroquinoline (EEDQ), N-isobutyloxy- carbonyl-2-isobutyloxy-1,2-dihydroquinoline (IIDQ). Preferred linkers may also be monomeric entities such as a single amino acid, especially preferred are those amino acids with small side chains, or a small polypeptide chain, or polymeric entities of several amino acids. Preferred polypeptide linkers are fifteen amino acids or less, more preferred are polypeptide linkers of ten or less amino acids. Even more preferred are polypeptide linkers of five or less amino acids. In an alternative embodiment, the linker may be a nucleic acid encoding a small polypeptide chain; preferred linkers encode a polypeptide of fifteen or less amino acids. More preferred linkers are nucleic acids encoding a small polypeptide chains of ten or less amino acids. Even more preferred linkers are nucleic acid encoding a small polypeptide of five or less amino acids, such as Gly-Phe-Leu-Gly, Gly-Gly, Gly-Leu or Gly, and the like.

Recombinant technology may be used to express a fusion MPP, linker and compound of interest, as described above and is well known in the art.

In another embodiment, the linker may be a cleavable linker, resulting in cleavage of the MPP and compound of interest once delivered to the tissue or cell of choice. In such an embodiment, the cell or tissue would have endogenous (either naturally occurring enzyme or be recombinantly engineered to express the enzyme) or have exogenous (e.g., by injection, absorption or the like) enzyme capable of cleaving the cleavable linker. Suitable enzymes for cleavage include, for example, use of a KEX2 protease recognition site (Lys, Arg) inserted between glucoamylase and the desired polypeptide to allow in vivo release of the desired polypeptide from the fusion protein as a result of the action of a native *Aspergillus* KEX2-like protease. (Conteras et al., 1991; Broekhuijsen et al., 1993; Ward et al., 1995). Another example of a cleavable linker peptide comprises the recognition sequence Asp-Asp-Asp-Asp-Lys, and wherein said fusion protein is cleavable by enterokinase.

Alternatively, the linker may be biodegradable such that the compound of interest is detached from the fusion MPP and compound of interest by hydrolysis and/or enzymatic cleavage inside cells. For example, tumors often express specific proteases, and be used in the delivery of prodrugs of cytotoxic agents. The linker may be selective for lysosomal proteases, such as cathepsin B, C, or D. Delivery of prodrugs and their subsequent activation is well

recognized, and such an approach provides significantly less systemic toxicity due to premature linker hydrolysis in the blood, consequently a greater amount of compound of interest, i.e., drug or cytotoxic agent, is delivered to the tumor site. See for example, T. Higuchi and V. Stella provide a thorough discussion of the prodrug concept in Pro-drugs as Novel Delivery Systems, Vol. 14 of the A.C.S. Symposium Series, American Chemical Society (1975). Examples of readily-cleavable groups include acetyl, trimethylacetyl, butanoyl, methyl succinoyl, t-butyl succinoyl, ethoxycarbonyl, methoxycarbonyl, benzoyl, 3-aminocyclohexylidenyl, and the like.

The compound of interest is any organic molecule, and includes small organic molecules, peptides, lipoproteins, and other modified proteins, polysaccharides, oligonucleotides, antisense oligonucleotides, and any other compound thought to have pharmaceutical, prophylactic, diagnostic properties and/or research interest. The compound of interest may be a small organic molecule already known to have pharmaceutical properties, and thus the present invention may be used as a method of treating a patient with the compound of interest. Alternatively, the compound of interest may be a novel protein of unknown function, and thus the present invention may be used as a method of identifying the function of the compound of interest. In another embodiment, the compound of interest may be an antisense molecule, and thus the present invention may be used as a method of altering transcription. In yet another embodiment, the compound of interest may be a prodrug, e.g. in an inactive form but capable of being activated once within the cell. In another embodiment, the compound of interest may be a cytotoxic agent, and thus the invention may be used as a method of delivering a cytotoxic agent to a cell. The compound of interest also includes detectable proteins which are useful to generate conjugated MMP and the detectable protein for identification of new MMPs. Detectable proteins include GFP, beta galactosidase, radiolabeled proteins and biotinylated proteins, proteins capable of conferring a detectable phenotype in the cell.

The present invention may be used to deliver the compound of interest into a cell *in vitro*, *ex vivo* or *in vivo*. For example, delivery may be carried out *in vitro* by adding the conjugated MPP and compound of interest extracellularly to cultured cells. Delivery may be carried out *ex vivo* by adding the conjugated MPP and compound of interest extracellularly or exogenously to a cultured sample removed from a patient, for example, blood, tissue or bone marrow, and returning the treated sample to the patient. Delivery may be carried out *in vivo* by administering the conjugated MPP and compound of interest by transdermal administration, inhalation, or injection to a patient.

Any type of cell may be used in the present invention. The cell may be of mammalian, bacterial, viral or yeast origin. The cell may be a cultured cell such as commonly used for oncology screening. Examples of cultured cells include CHO, HEK293T, HeLa, and NIH3T3. The cell may be a cultured cell from a patient suitable for *ex vivo* treatment with an MPP conjugate and reintroduction into a patient. The cell may be from the same or different patient than the patient to be treated.

Compositions of the invention comprising the conjugated MPP and compound of interest may be used for therapeutic, prophylactic, diagnostic or research purposes. Compositions may further comprise adjuvants, stabilizers and the like to improve the handling, stability and storage properties of the compositions.

Methods to identify novel MPPs are also part of the present invention. One method for identification of a membrane penetrating peptide is to generate a conjugate peptide comprising the sequence $-(X-X-X-X)_n-$ where n is an integer 1 to 7, and X each time is independently selected from the group consisting of arginine, histidine or lysine, with a detectable protein such as GFP, beta galactosidase and the like, adding the conjugate peptide to a cell and determining if the conjugated peptide is located within the cytoplasm and/or nucleus of the cell. Another method for identification of a membrane penetrating peptide is to generate a conjugate peptide comprising a peptide derived from or overlapping with a nuclear localization sequence of a mammalian or yeast protein and a detectable protein such as GFP, beta galactosidase and the like, adding the conjugate peptide to a cell and determining if the conjugated peptide is located within the cytoplasm and/or nucleus of the cell.

The following abbreviations are used for amino acids:

A refers to Ala, or alanine;

C refers to Cys or cysteine;

D refers to Asp or aspartic acid;

E refers to Glu or glutamic acid;

F refers to Phe or phenylalanine;

G refers to Gly or glycine;

H refers to His or histidine;

I refers to Ile or isoleucine;

K refers to Lys or lysine;

L refers to Leu or leucine;

M refers to Met or methionine;

N refers to Asn or asparagine;

P refers to Pro or proline;
 Q refers to Gln or glutamine;
 R refers to Arg or arginine;
 S refers to Ser or serine;
 5 T refers to Thr or threonine;
 V refers to Val or valine;
 W refers to Trp or tryptophan;
 Y refers to Tyr or tyrosine.

Proteins are written with the N-terminus to the left.

10 The following abbreviations are used: 'v/v' refers to volume to volume; 'EYFP' refers to a peptide fragment of the sequence Glu-Tyr-Phe-Pro; 'ORF' refers to Open Reading Frame; 'PCR' refers to polymerase chain reaction; 'CHO' refers to Chinese Hamster Ovary cells; 'HEK293T' refers to Human Embryonic Kidney cells, 'HeLa' refers to epithelial adenocarcinoma cells; 'NIH3T3' refers to Swiss mouse embryo fibroblast cells; 'DMSO' refers to dimethyl sulfoxide; 'FCS' refers to fetal calf serum; 'DMEM' refers to Dulbecco's Modified Eagle's Medium; 'PBS' refers to Phosphate buffered saline; 'BSA' refers to bovine serum albumin; 'C-terminus' refers to the carboxy-terminus; 'N-terminus' refers to the amino-terminus; 'PTD' refers to Peptide transduction domain; 'GPCR' refers to G-protein coupled receptor; 'TM' refers to a transmembrane domain of a GPCR; 'T' refers to an intracellular loop of a GPCR; '5HT2A' refers to serotonin receptor 2A; and 'mAb' refers to monoclonal antibody.

EXAMPLES

Example 1 Identification of an NLS within hPER1

Plasmid Construction

25 All hPer1 fragments described here are cloned as in-frame C-terminal fusion to EYFP. EYFP-hPer1 ORF, P1-N and P1-NX (fig. 1A) is generated by insertion of EcoRI and XhoI digested fragments into EYFP-C1 vector (Clontech). The other fragments are PCR amplified from the full-length hPer1 cDNA and subcloned into EYFP-C1 vector. The first and the last
 30 residue present in each of fragment is indicated in Fig. 1A. All constructs are verified by automated DNA sequencing.

Cell culture and DNA Transfection

CHO, HeLa and 293T cells are maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum (FCS), 50 units/ml penicillin, 50 μ g streptomycin, and 4 mM L-glutamine (hereafter referred to as complete DMEM) at 37°C with 5% CO₂. Transfection of the cells is carried in two-well Lab-Tek coverslips (Nunc Inc.) with LIPOFECT-AMINETMTM Reagent (Life Technologies) according to the manufacturer's instructions.

Peptides and Peptide Internalization

Peptides are synthesized by a commercial vendor (Bio Synthesis). For peptides internalization, cells are plated into two-well Lab-Tek coverslips (Nunc Inc.) at a density of 2×10^5 cells/well and cultured overnight. The peptides are dissolved in DMSO diluted to indicated concentration with PBS. The cell monolayers were incubated with the appropriate peptide/PBS solution at 1 μ M standard concentration for 10 min at room temperature (RT) unless otherwise specified. For experiments at 4°C, the protocol was the same except that all incubations were performed at 4°C until the end of the fixation procedure.

Immunofluorescence and Microscopy

For direct detection of expression and subcellular localization of EYFP fusion protein, transfected cells were examined directly without fixation or after fixation with 4%(v/v) formaldehyde in PBS for 20 min at 4°C and washed with PBS. For indirect immunodetection of biotinylated peptides, fixed cell were washed twice with PBS and permeabilized with 0.3% Triton X-100 in PBS for 20 min at 4°C and blocked with 2% BSA in PBS for 30 min at RT. Cells were then washed with PBS and incubated with Streptavidin-FITCTM (Sigma) or -Alex499 (Molecular Probe), 1:400 diluted in 0.2% Tween 20, 2% BSA in PBS for 1 h at RT. Following 2 x 5 min washes with PBS and once with 0.3% Triton X-100 in PBS for 20 min RT. In some experiment, the nucleus was stained with 50 ng/ml Hoechst 33258 (Sigma) or 3 μ g/ml propidium iodide in PBS. The subcellular localization of the fluorescence was analyzed on an Olympus microscope. Confocal images were taken on a Zeiss confocal laser scan microscope (CLSM phoibos 1000).

Though it is known that nuclear entry of PER1 is important for its function, no putative NLS was identified using a standard Profile Scanning program (Shearman, L.P., et al., *Neuron* **19**, 1261-1269 (1997), Yagita, K., et al., *Genes Dev.* **14**, 1353-1363 (2000)). To determine the NLS of hPER1 experimentally, three full-length hPER1 (P1-FL) were constructed and denoted as P1-N, P1-NM and P1-C (Fig. 1A). The ability of these constructs

to localize to the nucleus in CHO cells then analyzed. An EYFP-tag was used to facilitate the detection of hPER1 in living cells; however, the EYFP-tag had no apparent contribution on hPER1 fusion protein localization since hPER1 constructs made with an N-terminal Flag-tag presented an identical cytological distribution pattern (data not shown). After transient transfection, both P1-FL and P1-NM proteins were expressed in the nucleus of transfected cells as early as 10 hours post-transfection, while both P1-N and P1-C accumulated only in the cytoplasm (Fig.1B). The EYFP vector control was diffuse in both the nucleus and cytoplasm. These results demonstrate that a functional NLS in hPER1 is located between P1-N and P1-C in what we designated as region M (see Fig.1A).

To further localize the NLS in region M (amino acids 481-890), a series of 8 deletion constructs, P1-F1 to P1-F8, were generated and the subcellular distribution of each mutant was assayed as indicated in Fig. 1A and B. Sequential deletion from amino acid 581 (P1-F2) to position 821(P1-F7) of region M resulted in nuclear localization. Further deletion of amino acids 821 to 841 (P1-F8) resulted in a diffused fluorescent pattern within transfected cells with a localization pattern similar to that of the EYFP vector control. These data indicate that a NLS exists between amino acids 821 and 890, and is located at the C-terminus of region M. This observation was confirmed by the construction of an additional EYFP fusion protein, P1-NLS, which contained hPER1 amino acids 830-845. This region contains a string of basic residues that might function as a NLS (Weis, K., *Trends Biochem. Sci.* **23**, 185-189 (1998), Truant, R. and Cullen, B.R. *Mol Cell Biol.* **19**, 1210-1217 (1999)). As expected, P1-NLS exhibited nuclear localization in 100% of transfected cells (Fig. 1B). Other regions of PER1 in additional fusion constructs failed to localize to the nucleus (data not shown). Therefore, we conclude that the NLS of hPER1 (hPER1-NLS) is localized to within amino acids 830-845. Interestingly, construct P1-F1 has a strictly cytoplasmic localization pattern irrespective of the fact that it contains the NLS, supporting published observations that this region also contains and as yet unidentified cytoplasmic localization domain (Vielhaber, E., et al., *Mol Cell Biol.* **20**, 4888-4899 (2000)). Sequence alignment shows that the hPER1-NLS is conserved between human and mouse PER1 proteins (Fig. 1A), but not with other putative NLSs, or with other human, mouse or *Drosophila* PERs. After completion of our studies, Vielhaber et al. (2000), identified a longer mouse PER1-NLS that contains our identified 16 amino acid sequence (Vielhaber, E., et al., *Mol Cell Biol.* **20**, 4888-4899 (2000)); thus, supporting our findings.

Example 2 hPER1-NLS encodes an MPP

Two common features of the three identified gene encoded MPPs (TAT, Antp, and VP22) are that they are derived from nuclear proteins and they consist of basic amino acid residues (Lindgren, M., et al., *Trends Pharmacol Sci.* 3, 99-103 (2000)). hPER1 is also a nuclear protein whose NLS is rich in basic amino acids (SRRHHCRSKAKRSRHH, see Fig.

1). These similarities led us to determine whether hPER-NLS might also function as a MPP. In order to test this hypothesis, we synthesized several N-terminally biotinylated peptides: hPER1-MPP, Flag-tagged hPER1-MPP, Flag-tagged TAT-PTD, Flag-Flag alone, See Table 1 below:

Table 1

Name	Amino Acid Sequence	Transducing Peptide ¹	Nuclear Localizati on Fusion Protein ²
hPER1	GRRHHCRSKAKRSRHH	+	+
Flag-hPER1	GMDYKDDDDKGSRRHHCRSK AKRSHH	+	nd
Flag-TAT	GMDYKDDDDKGYGRKKKRR QRRR	+	+
Flag	GMDYKDDDDKG MDYKDDDDK	-	-
Antennapedia	GRQIKIWFQNRRMKWKK	+	nd
9 Arginine	GRRRRRRRRR	+	nd
9 Lysine	GKKKKKKKKK	+	nd
9 Histidine	GHHHHHHHHH	-	nd
NLSs:			
SV40	GDPKKRKRV	-	+
hPER2	GKKTGKNRKLKSKRVKPRD	-	+
hPER3	GRKGKHKRKKLP	+	+
Thyroid A-1	GKRVAKRKLIEQNRERRR	+	+

HME-1	GRKLKKKKNEKEDKRPRT	+	+
ABL-1	GKKTNLFSALIKKKKTA	+	+
Nucleoplasmin X	GRRERNKMAAAKCRNRRR	+	+
C-FOS	GRRERNKMAAAKCRNRRR	-	+
GCN-4	GKRARNTEAARRSRARKL	+	+
[R/H/K]-			
[R/H/K]-			
[R/H/K]-			
[R/H/K]			
HEN1/NSLC1	GRRRRATAKYRTAH	+	+
HEN2/NSLC2	GKRRRRATAKYRSAH	+	+
HNF3	GRRRRKRLSHRT	+	+
cAMP dependent TF	GRRRRRERNK	+	+
Cyclin L ania-6a	GKHRHERGHHRDRRER	-	+
beta Zip TF	GKKKRKLSNRESAKRSR	-	+
GFP		nd	-

Fn 1: Results shown for selected MPPs, see Fig 5

Fn 2: Results shown for selected MPPs, see Fig 5

The peptides are assayed for their ability to penetrate cellular membranes. Intracellular localization is assayed by direct staining with labeled Streptavidin ALEXA reagents or by indirect staining with anti-Flag mAb followed by the addition of labeled secondary antibodies. When added to the cells in culture at a concentration of 10 μ M, hPER1-MPP, Flag-hPER1-MPP and Flag TAT-PTD peptides are found to penetrate rapidly into 100% cells (Fig. 2A and Fig. 5). By both detection methods, hPER1-MPP, Flag-tagged hPER1-MPP, and Flag-tagged TAT-PTD are observed to be diffusely distributed throughout the cytoplasm, but concentrated within subnuclear domains that appear as distinct foci within the nucleoplasm and the nucleolus. In contrast, biotinylated negative control peptides, Flag-Flag and several additional peptides derived from other hPER1 regions are only barely discernible background staining, with no staining in the nucleus or nucleoli, even at high concentrations (data not shown). Confocal microscopy is used to confirm the intracellular and intranuclear staining of Flag-tagged hPER1-MPP, and that the negative control peptides are not internalized (Fig. 2A).

hPER1-MPP rapidly penetrated the cellular membranes and localized in nuclear regions with efficiencies similar to the TAT-PTD peptide (Fig. 2B). Identical results are obtained using CHO, HEK293T, HeLa, NIH3T3 and cultured rat primary cortical neurons (data not shown), indicating cell type-independent penetration.

hPER1-MPP internalization occurs rapidly (within 5 min), with similar potencies at 4°C and 37°C and even after cell membrane fixation (data not shown). Thus, the amino acid sequence 830-845 of hPER1 functions as both as a protein nuclear/nucleolar localization signal in the fusion protein and as a MPP, and that membrane penetration is independent of traditional receptor-mediated endocytic mechanisms.

Example 3 Arginine 7 is essential for hPER1-MPP activity

To date, the mechanisms as well as the structural basis whereby MPPs transverse cellular membranes have not been elucidated. We therefore sought to determine if there were key residues within hPER1-MPP that were important for maintaining those properties essential for its membrane penetrating potential. We separately replaced each amino acid in hPER1-MPP to alanine (Table 2), and assayed for the ability of these mutated peptides to penetrate living cells relative to the wild-type hPER1-MPP.

Alanine scanning:

Name	hPER1-PTD alanine substitution	Transducing Peptide
hPER1-PTD	SRRHHCRSKAKRSRHH	+
R2A	SARHHCRSKAKRSRHH	+
R3A	SRAHHCRSKAKRSRHH	+
H4A	SRRAHCRSKAKRSRHH	+
H5A	SRRHACRSKAKRSRHH	+
C6A	SRRHHARSKAKRSRHH	+
R7A	SRRHHCASKAKRSRHH	-
S8A	SRRHHCRASKAKRSRHH	+
K9A	SRRHHCRSAAKRSRHH	+
K11A	SRRHHCRSKAAARSRHH	+
R12A	SRRHHCRSKAKASRHH	+
S13A	SRRHHCRSKAKRARHH	+

R14A	SRRHHCRSKAKRSAHH	+
hPER1-PTD13	RRHHCRSKAKRSR	+
hPER1-Control (484-503)	QELSEQIHRLLQPV	-

As shown in Fig. 3, most of the single alanine substitutions had very little effect on membrane penetrating capabilities as compared with wild-type peptide. However, changing arginine 7 to an alanine (R7A) reduced the detectable cytological signal to that observed for the negative control peptides. Thus, the arginine 7 to alanine mutation significantly reduced the membrane penetrating properties of hPER1-MPP. Identical observations were observed after changing arginine 7 to glutamic acid (R7E) (data not shown). Furthermore, the simultaneous deletion of the N-terminal serine and of the two C-terminal histidines from hPER1-MPP (hPER1-MPP13) had little overall effect on the positive membrane penetrating potential of the peptide (Fig. 3).

The arginine 7 residue plays a critical role in the cell penetrating ability of the hPER1-MPP. We therefore sought to determine if the R7A mutation affected nuclear translocation of a fusion protein P1-NLS. CHO cells transfected with fusion protein P1-R7A (arginine 7 mutated to alanine) have intense nuclear staining similar to the wild-type, P1-NLS (data not shown). Nuclear translocation appears to be normal in the P1-R7A mutant fusion protein, but subnuclear targeting to the nucleoli is disrupted (data not shown). These data indicate that membrane penetration and nucleoli targeting are affected by the single R7 amino acid residue and indicate that nuclear translocation of hPER1-NLS has separate and distinct determinants.

Example 4 hPER1-MPP delivery of functioning molecules

One of the features of MPPs is their ability to cargo proteins and peptides into cells. We were successful in coupling hPER1-MPP to B-galactosidase and in delivering the fusion protein into cells in culture (data not shown), as has been described by Fawells et al., 1994 (Fawell, S., et al., *Proc Natl Acad Sci U S A*. **91**, 664-668 (1994)). However, to further extend the functional utility of MPPs, we tested hPER1-MPP in fusion with a physiologically relevant and biologically active peptide. Wess and colleagues (1993) have shown a functional role for the conserved transmembrane segment 7 (TM7) of the G-protein coupled receptor (GPCR) superfamily. Along with TM7, the third intracellular loop (I3) plays a significant role

in GPCR calcium signaling (Wess, JM., et al., *EMBO J.* **12**, 331-338 (1993)) while intracellular loops 1 and 2 (I1 and I2) appear not to be important. Using the serotonin receptor, 5HT2A, we experimentally tested the ability of hPER1-MPP and TAT-PTD in fusion with peptides designed from I1 and the TM7 domains to activate the receptor.

5 Biotinylated peptides hPER1-MPP TM7, TAT-PTD TM7, hPER1-MPP I1, TAT-PTD I1, hPER-MPP, TAT-PTD, TM7 or I1 were incubated at a concentration of 10 μ M with a 5HT2A receptor CHO stable cell line. Peptide membrane penetration was assayed using Streptavidin-Alexa 594 as described above. As shown in Fig. 4A, receptor signaling is activated by the addition of exogenous serotonin, hPER1-MPP TM7, and TAT-PTD TM7 as measured by the level of the calcium response. However, TM7 alone nor any of the other peptides were able to generate a calcium response. Furthermore, the activation of the receptor by hPER1-MPP TM7 and TAT-PTD TM7 is peptide concentration dependent, Fig. 4B. The addition of increasing concentrations of the activating peptide, TM7, in fusion with hPER1-MPP or TAT-PTD results in a calcium response in a dose dependent manner. TAT-PTD TM7 appears to be a more potent 5HT2A receptor activator than is hPER1-MPP TM7. A simple explanation for this result is that TAT-PTD TM7 is more cytoplasmically localized or has greater cell penetrating capabilities than hPER1-MPP TM7, although we have not observed that to be the case. Similar results were also obtained in this laboratory using hPER1-MPP in fusion with β -adrenergic activating peptides (unpublished data). These data support previous results that

10 hPER1-MPP not only penetrates cell membranes, but also demonstrates that it is capable of carrying peptides to intracellular compartments to initiate biologically relevant signal transduction events.

Example 5 Identification of other gene encoded MPPs

Since hPER1 is a nuclear protein proposed to be involved in transcriptional regulation, and since, to date, all PTDs derived from naturally occurring proteins are transcription factors (TAT, Antp, and VP22), we sought to determine if other PTD sequences existed within the genome. To this end, we used two approaches; first, we searched the NCBI non-redundant protein database for all known and putative NLS's (table 1, 10-17). We synthesized peptides corresponding to the NLS amino acid sequences and assayed for peptide transduction. As

25 shown in table 1 and Fig. 5, 6 of the 7 peptides synthesized had membrane penetrating characteristics similar to hPER-PTD and TAT-PTD. These proteins included human proteins of the thyroid hormone receptor alpha-1, homeobox protein HME1, and proto-oncogene protein ABL-1. Furthermore, (table 1 and Fig. 5) when we create in frame fusion proteins

between these peptide sequences and GFP then transfected into CHO or HEK 293T cells, all of the sequences conferred nuclear localization of the fusion protein.

Our second approach to identifying PTDs involved searching the NCBI non-redundant protein database collection with a degenerative algorithm (see Fig. 5, legend). Using these search parameters, we found 533,291 sequences of which the conditions for the algorithm were satisfied 129,169 times (24%). By limiting our search to include either “transcription factors, cytokines or tyrosine kinases”, we identified 8280 transcription factor protein sequences of which the algorithm pattern occurred 7374 times (89%); within 2333 cytokine protein sequences the pattern occurred 450 times (19%); and within 2513 tyrosine kinase protein sequences the pattern occurred 843 times (36%). Because the algorithm occurred most frequently in nuclear proteins, we synthesized peptides to putative PTDs for 6 of the “transcription factor” sequences and assayed for their ability to penetrate into the cells. As shown in table 1, results in lines 18-23 and fig 3A, 4 of the 6 peptides tested had membrane penetrating properties similar to hPER1-PTD and TAT-PTD. These proteins included two human proteins HEN1/NSCL-1 and HEN2/NSCL-2 which are reported to be involved in neuronal differentiation and development (Uittenbogaard, M., Peavy, D.R. and Chiamarello, A. 1999. Expression of the bHLH gene NSCL-1 suggests a role in regulation of cerebellar granule cell growth and differentiation. *J. Neurosci. Res.* **57**:770-781, Lipkowitz, S. et al. 1992. A comparative structural characterization of the human NSCL-1 and NSCL-2 genes. *J. Biol. Chem.* **267**:21065-21071), rat HNF-3 (17), and a *Drosophila* cAMP dependent transcription factor (18). Furthermore, (table 1 and Fig 5) when we create in frame fusion proteins between these peptides and GFP and transfected into CHO or HEK 293T cells, all of the sequences conferred nuclear localization of the fusion protein. These results indicate that PTD sequences can be found within or overlapping with NLSs. However not all NLSs are PTDs as is apparent in SV40, hPER2, C-FOS, Cyclin L ania-6 and beta Zip transcription factor NLSs (table 1). These results also suggest that PTDs sequences are prevalent throughout the genome and in particular within nuclear proteins.

Example 6. hPER-PTD with β -Galactosidase

At least one feature of HIV TAT transducing peptide is its ability to cargo proteins into cells and tissues. We therefore sought to determine if hPER1 transducing peptide could cargo beta galactosidase into cells. To perform this experiment, we followed a protocol by Frankel *et al.* PNAS 1989 (19):7397-401, whereby, we chemically linked hPER1-PTD or hPER-PTD

R7A(with Ala replacing Arg⁷) to full length β -galactosidase and assayed for the ability of these conjugates and beta-galactosidase protein alone to transduce into CHO cells. As shown in the figure 6, panel 1, cells incubated with hPER-PTD β -galactosidase fusion showed positive enzymatic activity for β -galactosidase as indicated by the blue color in the cells after the addition of X-gal. However, neither hPER-MPP R7A β -galactosidase nor β -galactosidase protein alone was able to enter the cells as indicated by a no blue staining reactivity after the addition of X-gal, panels 2 and 3. These data indicate that like TAT peptide, hPER1-PTD can cargo a large (120 kD) protein into cells.

001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053
054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2